

Estimation des paramètres des lois à queue régulière dans un cône convexe

Shuyan LIU

Université Paris 1 - SAMM
Université Catholique de Louvain - ISBA, Belgium

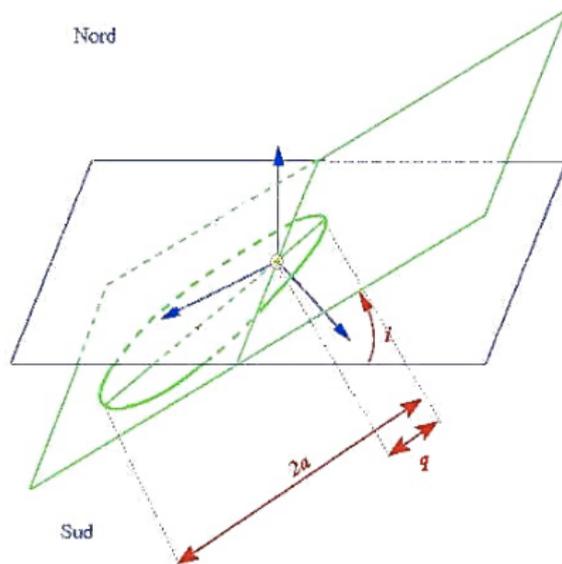
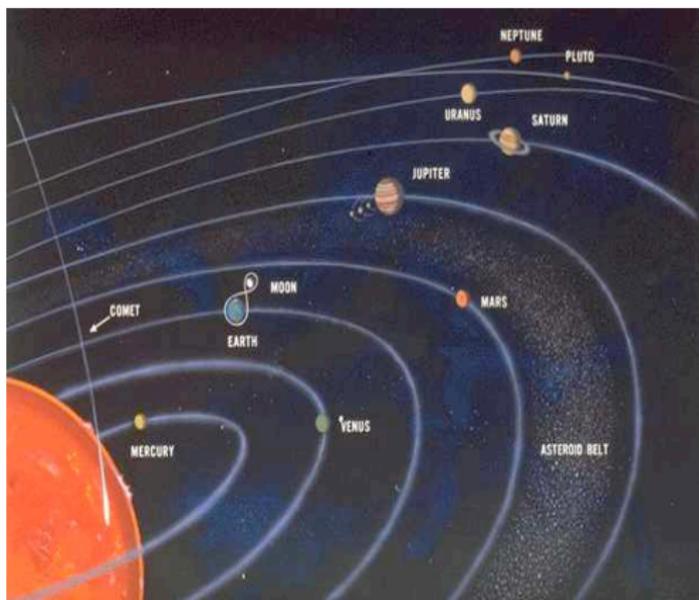
Youri DAVYDOV et Radu STOICA

Université Lille 1 - Laboratoire Paul Painlevé

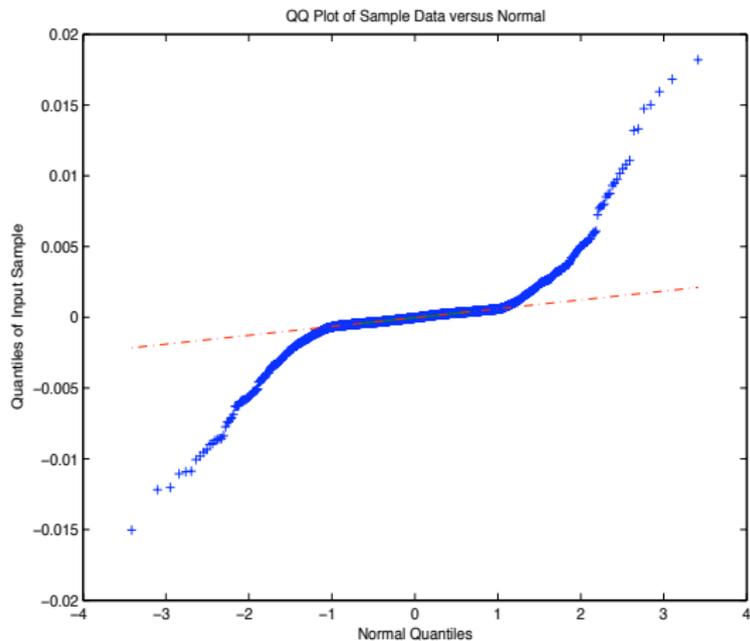
31 août-3 septembre 2010, Journées MAS et journée en l'honneur de
Jacques Neveu, Bordeaux

Un jeu de données des perturbations planétaires

Les données sont composées d'un ensemble des triplets $(\cos i, q, \Delta z)$,
 $z = 1/a$: l'inverse de demi-grand axe,
 $\Delta z = z_f - z_i$: la marque de perturbation.



QQ-plot des données dans une cellule autour de Jupiter

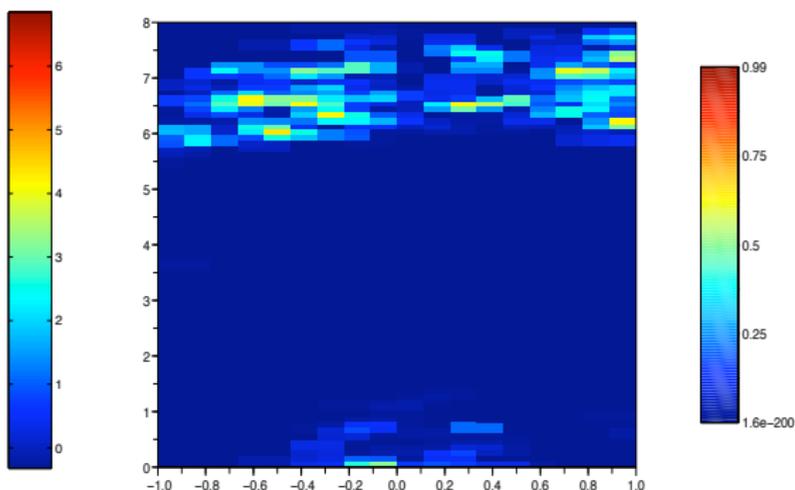
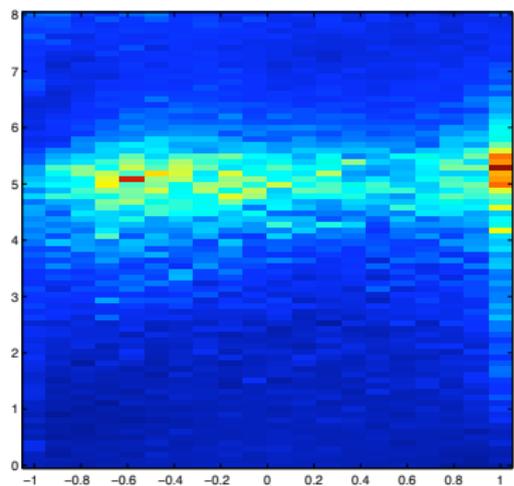


Comportement de type queue lourde autour des orbites des grandes planètes

Calculons dans chaque cellule l'**indicateur de queue lourde** :

$$\frac{\hat{z}_{0.99}}{\hat{h}_{0.99}} - 1.$$

Effectuons le test de normalité des quantiles empiriques $\hat{z}_{0.95}$.



Estimation des paramètres des lois à queue régulière dans un cône convexe

- 1 Définitions
- 2 Estimation des paramètres
- 3 Inférence statistique du modèle de lois à queue régulière
- 4 Conclusion et perspectives

Estimation des paramètres des lois à queue régulière dans un cône convexe

- 1 Définitions
- 2 Estimation des paramètres
- 3 Inférence statistique du modèle de lois à queue régulière
- 4 Conclusion et perspectives

Définition de lois à queue régulière

Un v.a. $X \in \mathbb{R}^d$ a une loi à queue régulière d'indice $\alpha > 0$ si $\exists \sigma$, une mesure finie sur $\mathbb{S}^{d-1} = \{x \mid \|x\| = 1, x \in \mathbb{R}^d\}$, telle que $\forall B \in \mathcal{B}(\mathbb{S}^{d-1})$ et $\sigma(\partial B) = 0$,

$$\lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > x \right\} = \sigma(B), \quad (X \in \text{VR}(\alpha, \sigma))$$

où L est une fonction à variation lente, i.e. $\frac{L(\lambda x)}{L(x)} \rightarrow 1$ quand $x \rightarrow \infty$, $\forall \lambda > 0$.

Une famille importante des lois à queue régulière est les lois non-gaussiennes stables

Un v.a. $X \in \mathbb{R}^d$ a une loi strictement α -stable si $\forall a, b > 0$

$$a^{1/\alpha} X_1 + b^{1/\alpha} X_2 \stackrel{\mathcal{L}}{=} (a + b)^{1/\alpha} X, \quad (\text{St}\alpha\text{S})$$

où X_1, X_2 sont des copies indépendantes de X .

Les lois α -stables n'ont pas d'expression explicite pour la densité

Exemple : Lois α -stables dans \mathbb{R}^1

Fonction caractéristique :

$$\mathbf{E} \exp itX = \begin{cases} \exp\{-\gamma^\alpha |t|^\alpha \left(1 - i\beta (\text{sign } t) \tan \frac{\pi\alpha}{2}\right) + i\mu t\}, & \alpha \neq 1, \\ \exp\{-\gamma |t| \left(1 + i\beta \frac{2}{\pi} (\text{sign } t) \ln |t|\right) + i\mu t\}, & \alpha = 1. \end{cases}$$

Paramètres :

$$\alpha \in (0, 2], \gamma = \sigma(1) + \sigma(-1) \in (0, \infty), \beta = \frac{\sigma(1) - \sigma(-1)}{\gamma} \in [-1, 1], \mu \in \mathbb{R}^1.$$

Exemples :

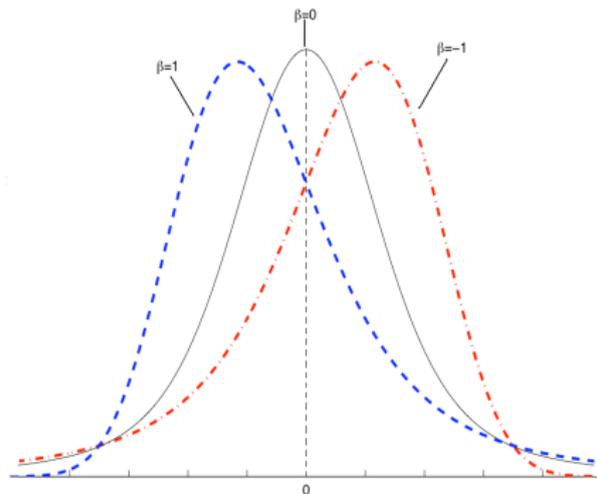
$$\alpha = 2 : \text{lois gaussiennes, } p(x) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{x^2}{4\sigma^2}}.$$

$$\alpha = 1 : \text{lois de Cauchy, } p(x) = \frac{\sigma}{\pi(x^2 + \sigma^2)}.$$

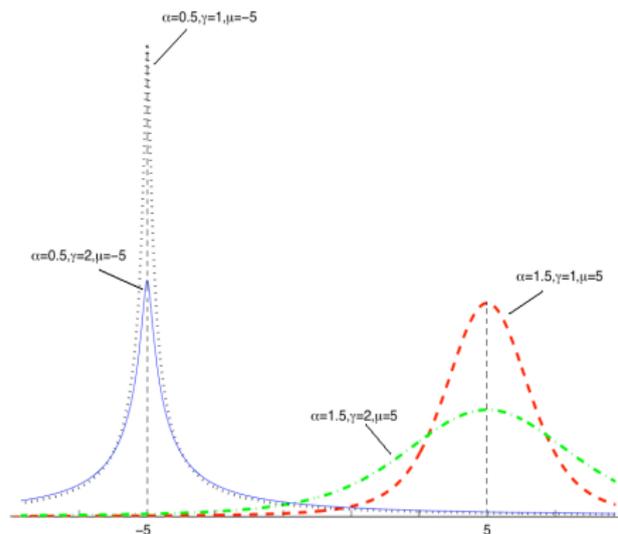
$$\alpha = 1/2 : \text{lois de Lévy, } p(x) = \left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{x^{3/2}} \exp\left(-\frac{\sigma}{2x}\right).$$

La densité de lois stables dans \mathbb{R}^1 dépend de quatre paramètres

Dépendance de β



Dépendance de α , γ et μ



La définition de lois stables a un sens dans un cône convexe

Un cône convexe \mathbb{K} est un semigroupe abélien topologique, supposé complet et séparable, avec une opération de multiplication par des scalaires positifs, i.e. $(x, a) \rightarrow ax$, continue pour $x \in \mathbb{K}$ et $a > 0$ tel que les conditions suivantes sont satisfaites:

$$a(x + y) = ax + ay, \quad a > 0, \quad x, y \in \mathbb{K}$$

$$a(bx) = (ab)x, \quad a, b > 0, \quad x \in \mathbb{K}$$

$$1x = x, \quad x \in \mathbb{K}$$

$$ae = e, \quad a > 0.$$

Avec différent choix du cône, on obtient différente famille de lois

Exemples :

- ▶ $(\mathbb{R}^d, +)$: lois α -stables.
- ▶ (\mathbb{R}_+^d, \vee) : lois max-stables.
- ▶ $([0, \infty), \vee)$: lois de Fréchet, $\Phi_\alpha = \exp(-(x/\sigma)^{-\alpha})$.
- ▶ $([0, \infty], \wedge)$: lois de Weibull, $F(x) = 1 - \exp(-(x/\sigma)^{-\alpha})$.
- ▶ (\mathcal{K}, \cup) : lois union-stables, où \mathcal{K} est une famille des sous-ensembles compacts de \mathbb{R}^d .

Estimation des paramètres des lois à queue régulière dans un cône convexe

- 1 Définitions
- 2 Estimation des paramètres
- 3 Inférence statistique du modèle de lois à queue régulière
- 4 Conclusion et perspectives

Algorithme d'estimation

Soit ξ_1, \dots, ξ_N des e.a. i.i.d. suivant une loi à queue régulière.

Étape 1 On divise l'échantillon en n groupes disjoints, chacun contient m éléments, i.e.

$$\underbrace{\xi_1, \dots, \xi_m}_{G_{m,1}}, \underbrace{\xi_{m+1}, \dots, \xi_{2m}}_{G_{m,2}}, \dots, \underbrace{\xi_{(n-1)m+1}, \dots, \xi_{nm}}_{G_{m,n}}$$

En pratique on choisit $n = \lfloor N^r \rfloor$ et $m = \lfloor N^{1-r} \rfloor$, $r \in (0, 1)$.

Étape 2 Notons $M_{m,i}^{(1)} = \max\{\|\xi\| : \xi \in G_{m,i}\}$,

$\xi_{m,i}$: l'élément dans $G_{m,i}$ tel que $\|\xi_{m,i}\| = M_{m,i}^{(1)}$,

$M_{m,i}^{(2)} = \max\{\|\xi\| : \xi \in G_{m,i} \setminus \{\xi_{m,i}\}\}$, $i = 1, \dots, n$.

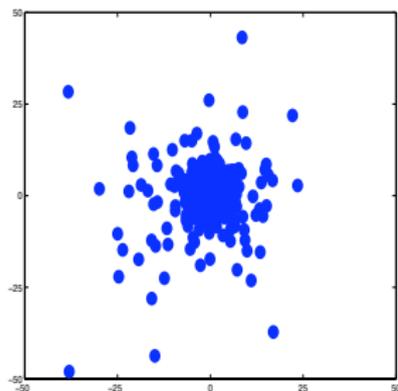
Étape 3 Calculons $S_n = \sum_{i=1}^n \frac{M_{m,i}^{(2)}}{M_{m,i}^{(1)}}$, $\theta_{m,i} = \frac{\xi_{m,i}}{\|\xi_{m,i}\|}$, $q_{m,i} = \frac{M_{m,i}^{(1)}}{m^{1/\alpha}}$,

$$\hat{\alpha}_N = \frac{S_n}{n - S_n}, \tag{2.1}$$

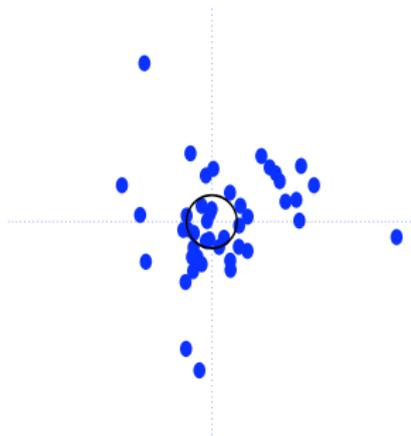
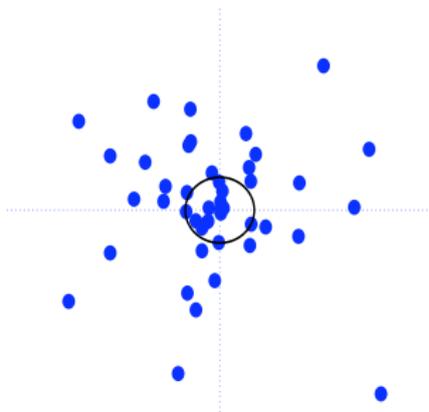
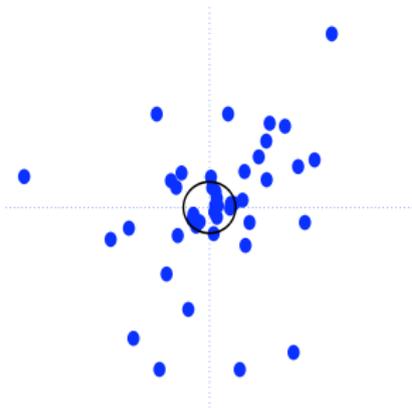
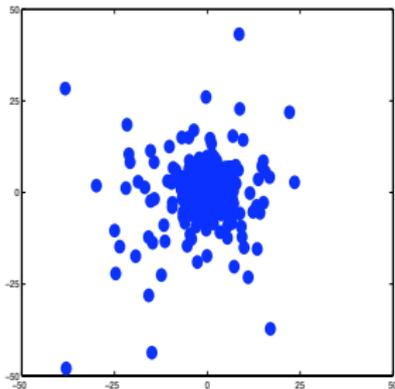
$$\hat{\sigma}_N(\cdot) = \frac{1}{n} \sum_{i=1}^n \delta_{\theta_{m,i}}(\cdot), \tag{2.2}$$

$$\widehat{\sigma(S)}_N = \left(\frac{1}{n\Gamma(1-\frac{\alpha}{t})} \sum_{i=1}^n q_{m,i}^t \right)^{\frac{\alpha}{t}}, \quad t > 0. \tag{2.3}$$

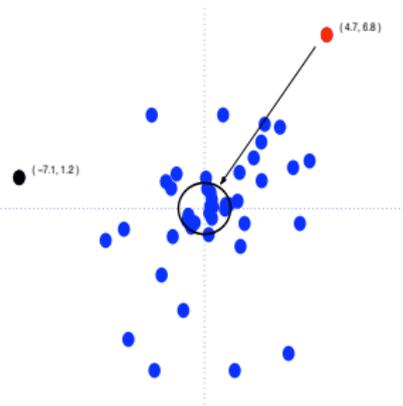
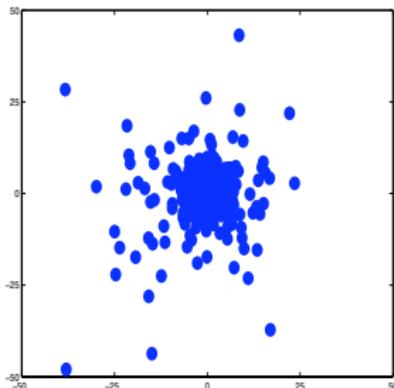
Un exemple de données simulées



Un exemple de données simulées



Un exemple de données simulées

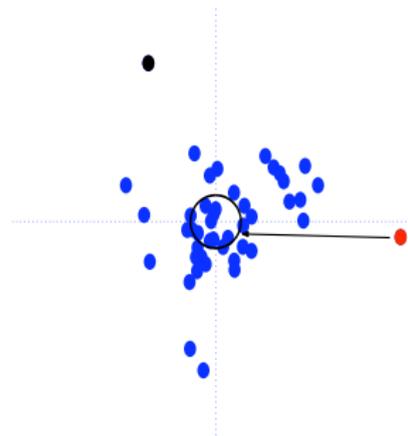
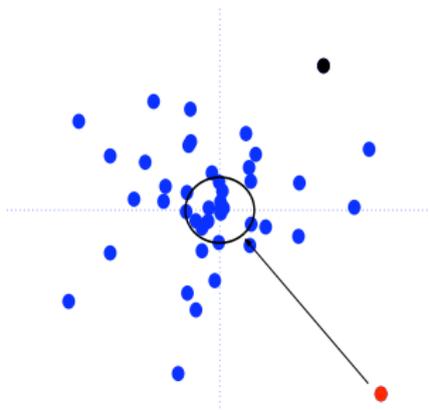


$$M_{m,1}^{(1)} = 8.3$$

$$M_{m,1}^{(2)} = 7.2$$

$$\frac{M_{m,1}^{(2)}}{M_{m,1}^{(1)}} = 0.87$$

$$\theta_{m,1} = \frac{(4.7, 6.8)}{M_{m,1}^{(1)}} = (0.57, 0.82)$$



$$S_n = \sum_{i=1}^n \frac{M_{m,i}^{(2)}}{M_{m,i}^{(1)}}$$

$$\theta_{m,i} = \frac{\xi_{m,i}}{\|\xi_{m,i}\|}$$

Estimateurs :

$$\hat{\alpha}_N = \frac{S_n}{n - S_n}$$

$$\hat{\sigma}_N(\cdot) = \frac{1}{n} \sum_{i=1}^n \delta_{\theta_{m,i}}(\cdot)$$

Consistance des estimateurs

Soient ξ, ξ_1, \dots, ξ_N des e.a. i.i.d. à valeurs dans \mathbb{K}' tels que $\xi \in \text{VR}(\alpha, \sigma)$.
Si S_n et $\hat{\sigma}_N$ sont défini par (2.1) et (2.2) avec $n \sim N^r$, $0 < r < 1$,
alors on a

$$\frac{1}{n} S_n \xrightarrow[N \rightarrow \infty]{p.s.} \frac{\alpha}{1 + \alpha}.$$

et

$$\hat{\sigma}_N \xrightarrow{p.s.} \sigma \text{ quand } N \rightarrow \infty.$$

Si la condition de régularité (VR) est satisfaite avec $L(x) = 1$ et $\widehat{\sigma(S)}_N$ est défini par (2.3) avec $n \sim N^r$, $0 < r < 1$, et $0 < t < \frac{\alpha r}{2}$, alors on a

$$\widehat{\sigma(S)}_N - \sigma(S) \xrightarrow[N \rightarrow \infty]{p.s.} 0.$$

Conditions pour établir la normalité asymptotique

- ▶ L'e.a. $\xi \in \mathbb{K}$ satisfait à **la relation asymptotique du second ordre** si

$$\mathbf{P}\{\|\xi\| > x\} = \sigma(S)x^{-\alpha} + cx^{-\rho} + o(x^{-\rho}) \text{ quand } x \rightarrow \infty, \quad (RS)$$

avec $0 < \alpha < \rho \leq \infty$ où c est une constante.

- ▶ L'e.a. $\xi \in \mathbb{K}$ satisfait à **la relation forte du second ordre** si $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$

$$\mathbf{P}\left\{\frac{\xi}{\|\xi\|} \in B, \|\xi\| > x\right\} = \sigma(B)x^{-\alpha} + cx^{-\rho} + o(x^{-\rho}) \text{ quand } x \rightarrow \infty, \quad (RFS)$$

avec $0 < \alpha < \rho \leq \infty$ où c est une constante.

Remarque

(RFS) $\Rightarrow \xi \in \text{VR}(\alpha, \sigma)$.

Théorème 2.4

Soit ξ un e.a. dans \mathbb{K}' tel que $\xi \in VR(\alpha, \sigma)$ et la condition (RS) est satisfaite.
Si on choisit

$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, \quad m = N^{1/(1+2\zeta)+\varepsilon},$$

où $\zeta = (\rho - \alpha)/\alpha$ et $\varepsilon > 0$, alors

$$\frac{\sqrt{n} \left(\frac{1}{n} S_n - \frac{\alpha}{\alpha+1} \right)}{\left(\frac{1}{n} \sum_{i=1}^n \left(\frac{M_{m,i}^{(2)}}{M_{m,i}^{(1)}} \right)^2 - \left(\frac{1}{n} S_n \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1).$$

Théorème 2.5

Soit ξ un e.a. dans IK' tel que la condition (RFS) est satisfaite.

Si on choisit

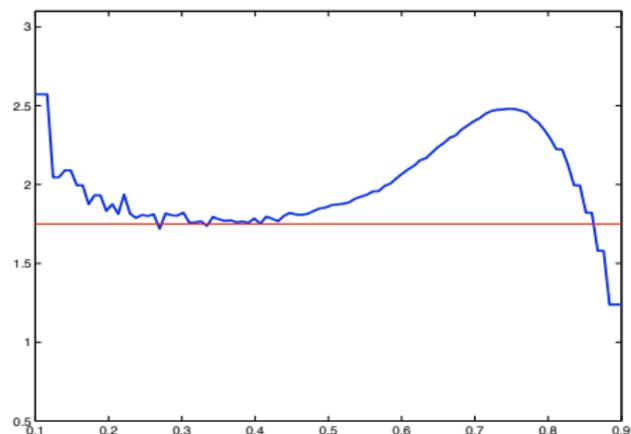
$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, m = N^{1/(1+2\zeta)+\varepsilon},$$

où $\zeta = \min(\frac{\rho-\alpha}{\alpha}, 1)$ et $\varepsilon > 0$, alors $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$ on a

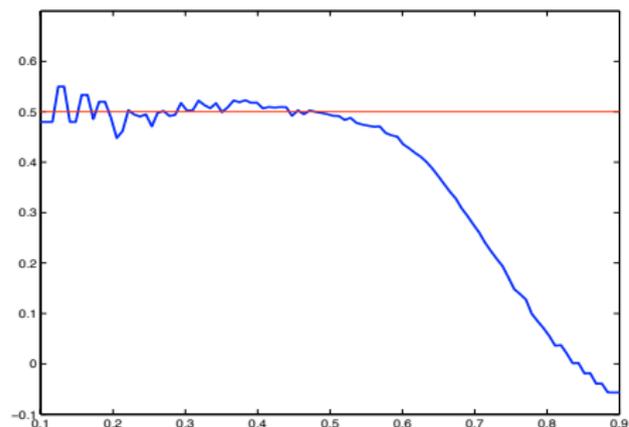
$$\frac{\sqrt{n}(\hat{\sigma}_N(B) - \sigma(B))}{\left(\frac{1}{n} \sum_{i=1}^n (\mathbb{1}_B(\theta_{m,i}))^2 - \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_B(\theta_{m,i}) \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1).$$

Influence du regroupement d'échantillon

Dans le cas de \mathbb{R}^d , le résultat de Fristedt implique $\rho = 2\alpha$. Cela nous permet de choisir $n = N^{2/3-\varepsilon}$, i.e. $r = 2/3 - \varepsilon$, alors la vitesse de convergence des estimateurs dans \mathcal{L}_1 s'approche de $N^{1/3}$.



$N = 100\,000, \alpha = 1.75$



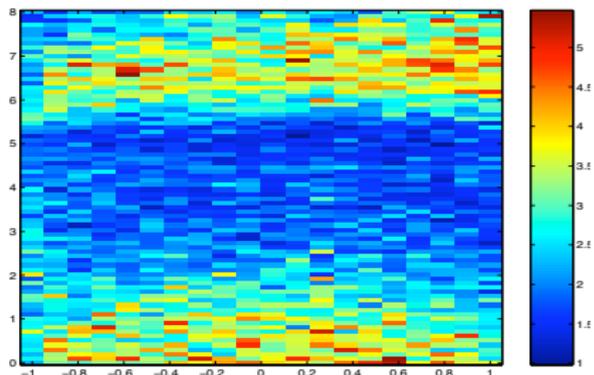
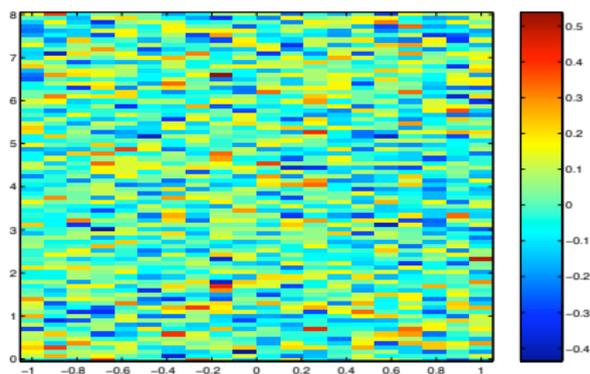
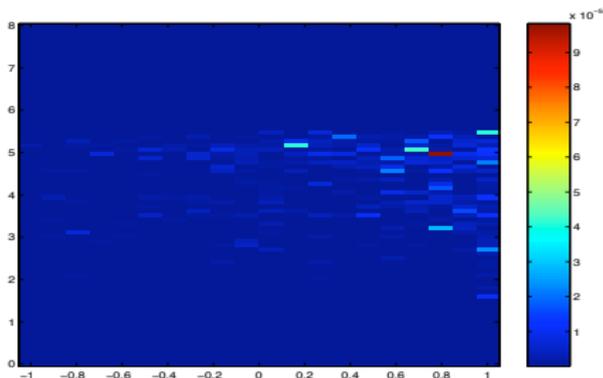
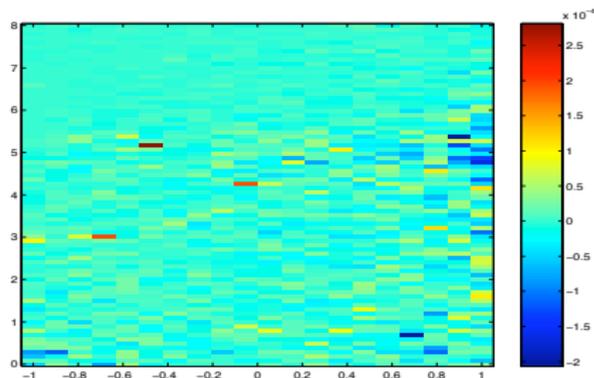
$N = 100\,000, \beta = 0.5$

Fig. Diagrammes des points $(1-r, \hat{\alpha}_N)$ et $(1-r, \hat{\beta}_N)$ estimés de loi $S(1.75, 0.5, 1)$.

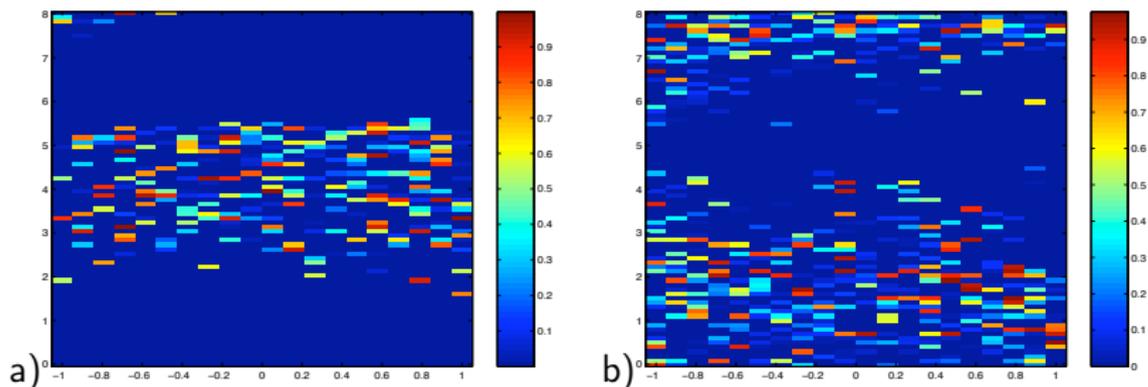
Estimation des paramètres des lois à queue régulière dans un cône convexe

- 1 Définitions
- 2 Estimation des paramètres
- 3 Inférence statistique du modèle de lois à queue régulière**
- 4 Conclusion et perspectives

Résultats d'estimation des paramètres pour les perturbations

 α  β  γ  μ 

Le modèle de lois à queue régulière donne un meilleur ajustement



a) $\alpha < 2$, H_0 : loi stable,

b) $\alpha > 2$, H_0 : loi alternative de densité définie par

$$f(z) = \frac{C_{\kappa, \alpha}}{1 + \kappa |z - \omega|^{\alpha+1}},$$

Estimation des paramètres des lois à queue régulière dans un cône convexe

- 1 Définitions
- 2 Estimation des paramètres
- 3 Inférence statistique du modèle de lois à queue régulière
- 4 Conclusion et perspectives

CONCLUSION

- ▶ Généralisation aux lois à queue régulière dans un cône abstrait
- ▶ Vitesse de convergence des estimateurs optimisée
- ▶ Modélisation des données réelles

PERSPECTIVES

- ▶ Estimateur de mesure spectrale : processus empirique indexé par des ensembles
- ▶ Estimation de la densité de mesure spectrale d'une loi stable
- ▶ La loi qui ajuste au mieux un jeu de données.
- ▶ ...

Merci!